Measuring & Predicting Visual Fidelity

Benjamin Watson

Dept. Computer Science
Northwestern University
watson@northwestern.edu

Alinda Friedman
Aaron McGaffey

Dept. Psychology
University of Alberta
alinda@ualberta.ca

The case for `int LooksLike()`

Models are often approximated
- *Model simplification*
- *Dynamic LOD*

Imagery is too
- *Image compression*
- *Image synthesis*
- *Video compression*

Existing stabs at `LooksLike()`

Among models
- *Distance*
- *Coplanarity*

In imagery
- *Mean squared error*
- *Models of human visual system*

Is `LooksLike()` working?

To begin, what do people think?

Some ways of finding this out:
- *Ratings* - (conscious)
- *Forced choice* - (conscious)
- *Naming times* - (subconscious)

Experiment: what people think

36 stimuli, subjects

Independent variables
- 2 simp methods: *QSlim, Cluster*
- 3 simp levels: *0%, 50%, 80%*
- 2 stimuli groups: *animals, objects*

Dependent variables
- *Ratings, preferences, naming*

One stimuli close-up

Unsimplified model

Simplified close-ups

50% simplified

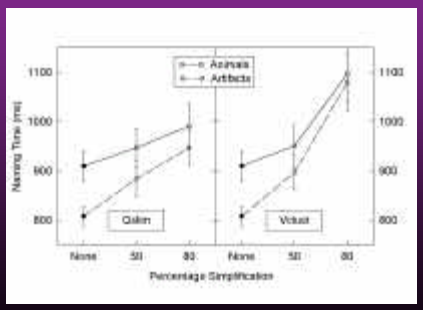
80% simplified

clustering qslim

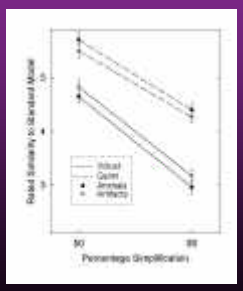
Animal stimuli

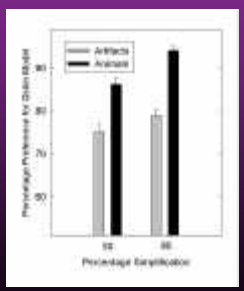
Artifact stimuli

Naming time results



Rating & choice results





Overall: what people think

	Simp Alg	Simp Level	Object Type
Naming Time	*QSlim*	*50%*	*Animals harder*
Ratings	*QSlim*	*50%*	
Forced Choice	*QSlim*	*80%: Q even better*	*Animals: Q even better*

Lessons:

Measures agree with intutition

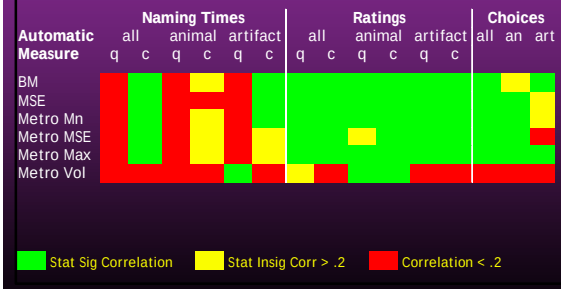
Measures disagree on object type

Is `LooksLike()` working?

Now, which `LooksLike()` to examine?

- *MSE:* image
- *BM:* image, perceptual [Bolin & Meyer]
- *Metro:* 3D [Cignoni, Rocchini & Scopigno]
  - Volume
  - Distance: mean, MSE, max

People vs. `lookslike()`



Limitations

One viewpoint

One fidelity manipulation

No background

No motion

No color

...

Confirmations

Results echo previous CHI study

Animal/artifact effect echoes psych

More simplification is worse

Qslim is better

Simplification harder at low poly counts

Surprises

Simplification success varies by obj type

- *Qslim best w/ animals, Clust w/ artifacts*

Differences in exp measures

- *Object type differences*

Naming/`LooksLike()` disagreement

- *Due to object type differences?*
- *Distillation effect?*

Implications

For simplification:

- *Specializations for model type?*
- *Small output is the real challenge*

For use of exp measures:

- *Ratings, choices: highly comparative*
- *Naming: more conceptual, subconscious*
- *How comparative is your app?*

Implications

For automatic measures:

- *MSE, BM, MetroMn all good*
- *Except, big naming problem!*

For future work:

- *Removing limitations*
- *Degree of comparison*
- *Naming & distillation effect*

Questions?